Quiz：错误分析

1. What are the differences between regression problem and classification problem? Accuracy: 89/99

•**Regression**: The objective is to predict a **continuous output**. This involves estimating a mapping function (f) from input variables (X) to a continuous output variable (Y). Regression is used to understand relationships between variables and for predicting values within a continuous range.

•**Classification**: The objective is to predict a **categorical output**. The task is to approximate a mapping function (f) from input variables (X) to discrete output variables (Y). The goal is to identify which category or class the input data belongs to.

Quiz：错误分析

2. For both regression problem and classification problem, what components are necessary for learning?

Accuracy: 34/99

1.Data
2. Feature Engineering
3. Model
4. Loss Function
5. Regularization Techniques
6. Learning Algorithm(gradient descent)

Get any four points

Quiz：错误分析

3. List all linear models you have learned so far?
4. List all nonlinear models you have learned so far?

Accuracy: 69/99

3. Linear regression, logistic regression, perceptron

4. KNN, decision tree, MLP

Quiz：错误分析

5. List all parametric models you have learned so far?
6. List all non-parametric models you have leaned so far?

Accuracy: 45/99

5. Linear regression, logistic regression, perceptron, MLP

6. KNN, decision tree

Quiz：错误分析

7. What are the differences between linear and nonlinear model? Accuracy: 90/99

- Decision Boundary:

Linear Models: The decision boundary in linear models is linear, meaning it can separatedifferent classes with a straight line (in two dimensions) or a plane (in three dimensions).Linear models are suitable for linearly separable datasets.

Nonlinear Models: Nonlinear models have a decision boundary that can be a curve or asurface, allowing them to adapt more complexly to the distribution of data. They are suitablefor datasets that are not linearly separable.

- Model Form:

Linear Models: The output of a linear model is a linear combination of the input features.expressed $y = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$ where $w_i$are the weights and $b$is thebias.

Nonlinear Models: The output of a nonlinear model is a nonlinear combination of the inputfeatures. which may include polynomial terms. exponential terms. logarithmic terms. etc. or introduces nonlinearity through activation functions, expressed as $y = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b)$, where f is a nonlinear function.

Quiz：错误分析

8. How does the decision boundary look like for all above models used for classification? Accuracy: 29/99

1. Logistic Regression
•Decision Boundary: Linear
2. Linear Regression (used for classification)
•Decision Boundary: Linear
3. Decision Trees
•Decision Boundary: Piecewise Linear and Axis-aligned
4. K-Nearest Neighbors (KNN)
•Decision Boundary: piecewise-linear, boundaries of Voronoi partition and Data-dependent
5. Perceptron
•Decision Boundary: Linear
6. Multilayer Perceptron (MLP)
•Decision Boundary: Highly Non-linear and Complex

Quiz：错误分析

9. What are the loss functions usually used for regression, binary classification, and multi-class classification, respectively? Write down the loss functions

Accuracy: 73/99

1. Regression

Mean Squared Error (MSE)

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Mean Absolute Error (MAE)

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

2. Binary Classification

Binary Cross-Entropy Loss

$$L(y, \hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

3. Multi-class Classification

Categorical Cross-Entropy Loss

$$L(y, \hat{y}) = -\sum_{i=1}^{n}\sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c})$$

Quiz：错误分析

10. What are the main differences for stochastic, mini-batch, and batch update? Accuracy: 80/99

Write down the MSE loss function for above three updates

Stochastic Gradient Descent (SGD)

In stochastic gradient descent, the model parameters are updated for **each training example**.

$$L(\theta) = \frac{1}{2}(y_i - f(x_i;\theta))^2$$

Mini-Batch Gradient Descent

Mini-batch gradient descent strikes a balance between stochastic and batch gradient descent by updating model parameters based on **a small subset** of the training dataset.

$$L(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - f(x_i;\theta))^2$$

Batch Gradient Descent

In batch gradient descent, the model parameters are updated based on the gradient of the loss function calculated over the **entire training dataset**.

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - f(x_i;\theta))^2$$

Quiz：错误分析

11. Write down the gradient-descent law and gradient-descent law with momentum, respectively.

Accuracy: 51/99

Gradient Descent Update Rule :

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t)$$

- θ represents the parameters of the model.
- L is the loss function.
- $\nabla_\theta L(\theta_t)$is the gradient of the loss function with respect to the parameters θ at iteration t.
- η is the learning rate, a scalar that determines the size of the step to take on each iteration.

Gradient Descent with Momentum

$$v_{t+1} = \gamma v_t + \eta \nabla_\theta L(\theta_t)$$

$$\theta_{t+1} = \theta_t - v_{t+1}$$

- v_t is the velocity at time step t.
- γ is the momentum coefficient.

Quiz：错误分析

12. How can we check overfitting? List all approaches for avoiding overfitting.

Accuracy: 48/99

Get any one points

Check overfitting:

- **Training Accuracy but Low Validation/Test Accuracy**: If the model performs exceptionally well on the training data but poorly on the validation/test data, it is likely overfitting.
- Plotting training and validation loss during training can show if the training **loss continues to decrease while the validation loss starts to increase**, indicating overfitting.
- **Cross-validation**

Get any four points

Approaches for avoiding overfitting:

- **Simplify the Model**
- **L1 and L2 Regularization**
- **Dropout**
- **Early Stopping**
- **Data Augmentation**
- **Increase Training Data**

Quiz：错误分析

13. What are training set, validation set, and test set used for?

Accuracy: 95/99

1. Training Set
The training set is used **to train the model**. This set is used to **fit the model's parameters**.

2. Validation Set
The validation set is used to provide an **unbiased evaluation of a model** fit on the training dataset while **tuning the model's hyperparameters** (like learning rate, number of layers, etc.). This set is crucial for **preventing the model from overfitting.**

3. Test Set
The test set is used to provide an unbiased **evaluation of a final model** fit on the training dataset. The test set **evaluates the model's performance to generalize to new, unseen data**.

Quiz：错误分析

14. Draw a 3-layer MLP network for 3-class classification, where the input features have dimension 4, and hidden layers both have 5 units

Accuracy: 94/99